Cell Stem Cell
Forum

Whole-Genome Analysis, Stem Cell Research, and the Future of Biobanks

Thane Kreiner[1,*] and Stefan Irion[1]
[1]The Center for Science, Technology, and Society, Santa Clara University, Santa Clara, CA 95053 USA
*Correspondence: tkreiner@scu.edu
http://dx.doi.org/10.1016/j.stem.2013.04.024

The convergence of the genomic revolution and biobanking with rapid progress in stem cell research holds vast potential for personalized medicine and novel therapies. In this Forum, we explore social and ethical considerations emerging from strategies to realize the promise of these remarkable technologies.

In 2005, one of the first genome-wide association studies (GWAS) was published, in which over 116,000 SNPs in 146 individuals were screened to reveal a complement factor H polymorphism associated with age-related macular degeneration (Klein et al., 2005). Some 2 years later, successful reprogramming of human fibroblasts to induced pluripotent stem cells (iPSCs) was reported (Takahashi et al., 2007).

Whole-genome analysis technologies, including microarrays and next-generation sequencing (NGS), are fueling the discovery of genetic variants associated with a wide variety of human conditions; both effect size and allele frequency are critical dimensions in associating genetic variation with diseases and other human conditions (Bush and Moore, 2012). Genetic variants explain only a portion of phenotypes, with environmental exposures and epigenetic factors influencing gene expression and ultimately, human traits. While stem cell technologies, including human embryonic stem cells (hESCs), iPSCs, transdifferentiation, dedifferentiation (Jopling et al., 2011), and adult stem cells are rapidly accelerating our understanding of biology and disease, none of these have been applied to genetically characterized samples from a broad population base.

The convergence of next-generation stem cell and whole-genome analysis technologies presents an opportunity to create disease-relevant cell types with known genetic variants that represent human populations and could accelerate personalized medicine, advance drug discovery, and broaden our understanding of disease biology. Such an approach would discern the biological repercussions of the alleles, whether common or rare. The consequent knowledge could be leveraged to find new therapeutics and design smarter clinical trials with inclusion and exclusion criteria tailored to expected responses based on genetic variation (Bellin et al., 2012). Moreover, biobanks of cells with known and defined genetic profiles could facilitate development of cell-based therapeutics less prone to immune rejection. Researchers, funding agencies, and policy makers will need to converge on a common set of biological, ethical, and social principles.

Biobanks as a Potential Source of Next-Generation Stem Cells

To fuel the next wave of regenerative medicine research and clinical application development, integrated biobanks are essential. These integrated biobanks would support GWAS and also serve as source material for next-generation stem cells technologies with known and specific genetic composition.

Biobanks are repositories of samples for use in research studies. The biological material itself is practically useless without some associated information. Age, sex, and continent(s) of origin are a few attributes nearly universal to human biobanks. Biobanks often include medical information, particularly if they have been constructed to accelerate research in a particular disease area. Some capture longitudinal information from individuals to support population studies, others are disease-specific, and still others comprise residual tissues from medical procedures or diagnoses (Knoppers et al., 2012).

Large-scale biobanks that support population-based studies such as GWAS have raised numerous ethical, legal, and social issues, among them the potential identification of donors, revelation of individual genetic predispositions, and information about biological family members who may not have provided any type of informed consent (Cambon-Thomsen, 2004; Knoppers et al., 2012). Moreover, there is significant variation among informed consent for biobanks: specific to a particular study or broad; obligations, if any, to return results to donors; and the right of donors to withdrawal from future studies (Knoppers et al., 2012). These issues persist and some are potentiated by the convergence of large-scale genomic studies and next-generation stem cell technologies.

However, stem cell experiments differ conceptually from whole-genome analysis: they are based on individual samples rather than populations. Researchers generally use readily available cellular material (e.g., rodent cells, cell lines, or opportunistic clinical specimens) that may not have been genetically characterized. Well-concerted efforts to link stem cell generation and biobanks of representative human populations are in their infancy. One fundamental driver is that biobanked samples for stem cell generation have a high bar for preservation: living cells. Existing tissue biobanks, however, often lack live, recoverable cellular material because the collections predate next-generation stem cell technologies.

Each of these next-generation technologies presents challenges with respect to biobank creation. It is unknown whether somatic cells commonly used in reprogramming such as dermal fibroblasts retain cell-of-origin epigenetic memory (Sullivan et al., 2010) or whether epigenetic markers will affect models of disease in derived cell lineages for research or cell-based therapy. In contrast, dedifferentiated cells may be epigenetically close to

the cell they are destined to resemble. Here, only a limited number of steps are necessary to elicit proliferation without reverting all the way to pluripotency. Adult stem cells are local, organ-specific stem cells and may be more epigenetically similar to their differentiated counterparts within the organ but are extremely rare in most tissues. Since dedifferentiated and adult stem cells require the use of local resident cells, a very limited number of biobanks exist. Creating new banks for these methods would require the collection of cells that are not readily accessible, such as neurons or pancreatic islet cells, presenting a significant obstacle for some conditions. Furthermore, suitable long-term storage methods for some cell types do not currently exist.

Future biobanked samples for any stem cell approach should include accompanying clinical, biological, and sequence information. Appropriate informed consent will also be necessary and must be considered in the broader context of the accentuated ethical, legal, and social issues emerging from the convergence of whole-genome analysis and next-generation stem cell technologies.

Ethical, Social, and Policy Issues

Even before the formal launch of the Human Genome Project (HGP) in 1990, the US National Institutes of Health formed a joint working group with the Department of Energy to "identify and address" the ethical, legal, social, and economic issues (ELSI) that would arise with technology advancements for genome analysis (see Web Resources). The first 5 year plan for the HGP included an ELSI budget allocation, which continued throughout the 13 year project. Among these issues are: privacy and confidentiality, the potential for discrimination based on genetic information, and intellectual property.

The HGP and programs such as Small Business Innovation Research (SBIR) and Advanced Technology Program (ATP) grants accelerated development of genome analysis technologies in the private sector. By the time the HGP was declared complete in April 2003, a number of whole-genome analysis studies were well underway. These advances in genome analysis technology quality and declines in cost have made privacy protection for biobank samples untenable: recently, the ability to recover surnames from personal genomes and link them to deidentified public data sets was demonstrated (Gymrek et al., 2013). As direct-to-consumer genetic tests and personal genome initiatives proliferate, vast amounts of genetic information are being generated, stored, and searched. DNA from shed skin cells or hair can be collected without permission or knowledge; massive law enforcement databases, such as CODIS (see Web Resources), already exist. Affordable, easy-to-use DNA sequencing technology will almost certainly become widely accessible.

Biobanked tissues and cells contain the full genomes of the donors; thus, they include the same ELSI concerns as DNA samples. Application of next-generation stem cell technologies to existing biobanks will foster applications beyond the original intent of the biobank. New cells, cell lines, or tissues with known genetic profiles are a likely outcome of the convergence between genome analysis and stem cell technologies. Such products may outlive the intended research objectives by several decades, raising new ELSI concerns. Nonautologous cells or tissues generated from iPSCs or dedifferentiated cells for regenerative medicine applications carry the donor's genetic information, subverting the notion of autonomy or control over one's own genetic information; in theory, the recipient could decipher the donor's entire genetic code, revealing alleles that correlate with certain behaviors or early-onset, untreatable diseases.

Protection against genetic discrimination remains a vital ELSI concern. However, it was not until 2008 that the Genetic Information Nondiscrimination Act finally passed the US Congress; its protections are limited to employment and health insurance. In 1997, UNESCO adopted the Universal Declaration on the Human Genome and Human Rights (see Web Resources), Article 6 of which provided a more comprehensive statement against genetic discrimination, including preservation of "human rights, fundamental freedoms, and human dignity." Converging advances in genetic analysis and regenerative medicine magnify the need for broad genetic information nondiscrimination legislation in UN member nations. Genetic information should not provide a basis for discrimination in access to social services, educational opportunities, or other aspects of living. Donors to integrated biobanks should be assured that their participation in no way jeopardizes future opportunities for themselves or biological family members.

Informed consent principles for biobanked samples are particularly complex as whole-genome analysis and regenerative medicine technologies converge. An underlying principle of informed consent is that individuals choose how and whether their tissues, cells, DNA, or associated data will be used. Some scholars limit the definition of true informed consent to specific, well-defined research; others propose that broader consent is a tractable solution for longitudinal population studies. The premise that research studies be approved by institutional or ethics review boards is universally held.

The future use of samples presents a fundamental informed consent challenge that is accentuated by the preservation of living tissue that can, in theory, generate genetically identical tissues, organs, or organisms that may outlive both the donor and the researcher. Because science is evolutionary and combinatorial, the design of future studies cannot be anticipated at the time samples are collected. Notification, opt-out, withdrawal, return-of-results, and incidental finding provisions provide some potential solutions (Cambon-Thomsen, 2004; Knoppers et al., 2012); however, they impose research transaction costs, which may appear virtually unlimited in the case of incidental findings. Implementation of some informed consent provisions after death may unintentionally compromise the privacy of survivors.

Integrated informed consent will be an essential feature of next-generation biobanks. Since the banks will contain live cells with full genetic information, consent must encompass future uses of the material and the knowledge coming from it. Engineered tissues could replace long-term pharmaceutical treatment for chronic conditions such as diabetes and heart disease; would donors of the progenitor cells be informed of results in recipients, compromising their privacy? Would recipients have a right to know the genetic code of tissues implanted in their own bodies? Notification and opt-out provisions could become particularly problematic. Would initial cell donors or their survivors have any rights to preclude use

of derived tissues in certain individuals or populations because of religious or other personal beliefs? All future uses cannot be anticipated or imagined; therefore, ethics review committees proficient in both genetic analysis and stem cell research will be vital, raising the need for integrated professional development. In addition to considering the ELSI implications of genetic information and regenerative medicine, a "tiered" informed consent approach might enable donors to opt out of future, undefined applications.

Although a meaningful review of gene and stem cell patents is beyond the confines of this Forum article scope and space, ownership of biobank samples and commercial use provisions related to intellectual property rights merit some remark (for further reading see Mathews et al., 2013). Soon, the US Supreme Court will hear the *Association for Molecular Pathology v. Myriad Genetics* case challenging the validity of the *BRCA1* and *BRCA2* gene patents in the US; whether naturally occurring gene sequences are patentable varies by country. Similarly, isolated stem cells are patentable in the US, but the European Court of Justice ruled against patents on hESC lines (see Web Resources).

Patent claims may relate to composition of matter, e.g., the DNA sequence or stem cell itself, or to the method or process for using matter. Patentable subject matter usually needs to meet three tests: novel, nonobvious to one skilled in the art, and useful. Composition of matter patents issued on naturally occurring gene sequences claim novelty because the sequence is isolated; once a single gene had been isolated and sequenced, it seems obvious that one would wish to isolate and sequence all of them; hence, the Human Genome Project. The patentability of technologies for genome analysis, however, is broadly accepted.

Patent filings for pluripotent stem cell generation and directed differentiation abound. Because these methods are relatively novel, it remains to be seen whether dominant technology claims will emerge such as Cohen-Boyer patent for recombinant DNA technology. Many of the earliest processes using viral vectors are unlikely to yield cells or tissues that are safe for regenerative medicine applications; transient gene suppression for dedifferentiation or reprogramming followed by proliferation may have more clinical utility. In any case, the recent, rapid stem cell research progress suggests that useful applications will emerge and require meaningful intellectual property protection.

From an intellectual property perspective, the convergence of whole-genome analysis and stem cell technologies raises and accentuates challenging issues: who "owns" an individual's genetic information and the biology it encodes? What rights, if any, should donors have to profits generated from organs or tissues derived from their cells? Who decides on limitations for use of DNA, cells, tissues, or organs? The courts thus far have held that donors do not have ownership rights in biological materials derived from their samples; however, the value of genetically proficient replacement hearts, livers, and islet cells could impact the calculus of commercialization over time, especially as individuals are empowered to understand their own genomes through consumer genetics services.

The ethical and legal issues for regenerative medicine diminish substantially when new stem cell technologies are employed that enable use of the patient's own cells as starting material; the costs and time to cell and tissue therapies become predominant factors in the viability of these approaches. However, it may not be advisable to treat some medical conditions with autologous regenerated cells: they will bear the same genetics, and possibly, the same epigenetics. Thus, while autologous cells may circumvent some ELSI concerns, biobanking of allogeneic, genetically profiled cells may be necessary to realize personalized and affordable regenerative medicine. The ELSI considerations are profound and warrant ongoing investment to fuel thoughtful analysis.

Public investment is essential to realize the promise of this technology convergence. Integrated biobanks that support the convergence of GWAS and stem cell research require infrastructure and standardization. No perfect technology to generate stem cells at scale has yet emerged. Viral reprogramming is best established, but raises safety concerns. Adherent fibroblasts are more suited for automation, but lymphocytes are more easily collected. Optimization of methods and comparative analysis is the first step toward standardization; such efforts are underway at several institutions and in industry. To limit variability, stem cell lines must then be generated using standard operating procedures (SOPs). Moving from laboratory-scale stem cell generation to large cohort analysis raises new challenges. First, how many clones need to be generated and characterized per donor? Investigators using iPSCs often include three to five clones per individual, which would catapult a cohort of 300 donors to over 1,000 samples. Second, how many donors represent a population (a challenge familiar to GWAS)? Third, should donors be followed over time to collect material prior to, during, and after disease to capture potential epigenetic changes? While costs are declining, stem cell generation and differentiation remains laborious and expensive. A smart selection based on allele effect size may help to reduce these numbers. Once integrated biobanks are established, standard methods to differentiate and interrogate cells of interest also need to be developed. Funding must be adequate to solve these technical, biological, and ethical scenarios over time. Granting agencies have already crafted RFAs designed to surface and address some of these infrastructure and standardization issues (see Web Resources).

Future Directions

We explore the convergence of two seemingly unrelated technologies that now can act in concert to advance our understanding of human disease and accelerate development of potential therapies. To realize this potential, we need to reimagine biobanks as core infrastructure to generate stem cells on a population scale. Future "centers of excellence" for biobanking should employ cell biologists, bioinformaticians, clinicians, ethics advisors, and policy makers. At these centers, standardized cell collection and cell production should be driven by SOPs; pertinent biomedical and bioinformatics data should be available in an open, searchable format with the option and requirement for resubmission of phenotypic and cellular data.

Both the biobank infrastructure and thoughtful analysis of novel as well as accentuated ethical, legal, and social issues emerging from the convergence of

whole-genome analysis and stem cell research require significant investment. Some initial efforts are underway through CIRM and NIH (see Web Resources); recently, the Japanese government announced a $1 billion investment in regenerative medicine ventures based on iPSCs (see Web Resources). While difficult to quantify, platform investments in biological sciences seem to pay off: the US government estimates an economic impact of $140 for each of the $3.8 billion invested in the HGP from 1990–2003 (see Web Resources). Imagine the collective improvement in quality of life and economic return with investment that accelerates the convergence of genome-wide analysis and stem cell research.

WEB RESOURCES

The URLs for data presented herein are as follows:

Embryonic Stem Cells. http://stemcells.nih.gov/Pages/Default.aspx

Nobel Prize for cellular reprogramming http://www.nobelprize.org/nobel_prizes/medicine/laureates/2012

Stem cells clinical trials. http://clinicaltrials.gov/ct2/show/NCT01217008?term=geron&recr=Open&rank=5

CODIS. http://www.fbi.gov/about-us/lab/biometric-analysis/codis

Human Genome Project ELSI issues. http://www.ornl.gov/sci/techresources/Human_Genome/publicat/hgn/v2n1/05elsi.shtml

European hESCs. http://www.nature.com/news/2011/111024/full/478441a.html http://www.ornl.gov/sci/techresources/Human_Genome/publicat/hgn/v2n1/05elsi.shtml

RFAs to standardize stem cell production. http://www.cirm.ca.gov/about-cirm/newsroom/press-releases/03192013/stem-cell-agency-banks-32-million-new-approach-advance

UNESCO Declaration. http://portal.unesco.org/en/ev.php-URL_ID=13177&URL_DO=DO_TOPIC&URL_SECTION=201.html

Japan's investment in iPSCs. http://www.economist.com/news/business/21572235-best-market-world-right-now-regenerative-medicine

Return on the Human Genome Project. http://www.genome.gov/27544383

REFERENCES

Bellin, M., Marchetto, M.C., Gage, F.H., and Mummery, C.L. (2012). Nat. Rev. Mol. Cell Biol. *13*, 713–726.

Bush, W.S., and Moore, J.H. (2012). PLoS Comput. Biol. *8*, e1002822.

Cambon-Thomsen, A. (2004). Nat. Rev. Genet. *5*, 866–873.

Gymrek, M., McGuire, A.L., Golan, D., Halperin, E., and Erlich, Y. (2013). Science *339*, 321–324.

Jopling, C., Boue, S., and Izpisua Belmonte, J.C. (2011). Nat. Rev. Mol. Cell Biol. *12*, 79–89.

Klein, R.J., Zeiss, C., Chew, E.Y., Tsai, J.Y., Sackler, R.S., Haynes, C., Henning, A.K., SanGiovanni, J.P., Mane, S.M., Mayne, S.T., et al. (2005). Science *308*, 385–389.

Knoppers, B.M., Zawati, M.H., and Kirby, E.S. (2012). Annu. Rev. Genomics Hum. Genet. *13*, 395–413.

Mathews, D.J.H., Cook-Deegan, R., and Bubela, T. (2013). Cell Stem Cell *12*, this issue, 508–512.

Sullivan, G.J., Bai, Y., Fletcher, J., and Wilmut, I. (2010). Mol. Hum. Reprod. *16*, 880–885.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Cell *131*, 861–872.